Against the Grain

Volume 19 | Issue 1 Article 8

February 2007

Zen and the Digital Collection Librarian

James A. Bradley
Ball State University, jabradley2@bsu.edu

Zen and the Digital Collection Librarian

by **James A. Bradley** (Head of Metadata and Digital Initiatives, Ball State University, University Libraries, BL-025, Ball State University, Muncie, IN 47306: Phone: 765-285-5718) <jabradley2@bsu.edu>

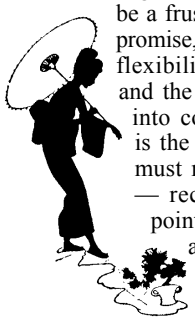
"The container tends to shape the contained."

It sounds a bit like an eastern proverb — and, to be completely honest, I'm not certain that I haven't unconsciously borrowed the phrase from one of the many poorly-dubbed Kung Fu films I indulge in from time to time.

Whatever the origins, the concept is simple: Flexible matter immediately assumes the shape of whatever you pour it into; and even a somewhat rigid object will, over time, succumb to the contours of its packaging. By the same token, a rigid item will simply break if forced into a container that is too foreign or restrictive.

I've found this maxim to be a true and useful analogy in the planning of digital collections.

Bringing an existing "real world" collection, and its accompanying metadata, across the digital threshold can sometimes be a frustrating process — full of promise, but also compromise. The flexibility of both the container and the contained must be taken into consideration; however, it is the collections librarian who must remain the most willowy — recognizing the shattering points of both and finding an appropriate fit.



Such was the case of the **Ball State University Architecture Image Collection**.

The challenge to **Ball State University Library**'s new Digital Initiatives program was to migrate the visual resources of the Architecture Library into a single online environment that would facilitate remote access, advanced searching capabilities, and image delivery at a resolution suitable for research and classroom instruction within the College of Architecture and Planning (CAP).

The first step in the conversion process was to assess and gather the characteristics of the materials to be digitized:

- Approximately 120,000 35mm slides
- Local call number for access purposes
- Group level **MARC** records that gather individual slides according to location or site.

The next step was to consider the characteristics of the desired online collection:

- Slides must be scanned and stored in accordance with archival standards.
- Derivative images must be created for online delivery.
- The "front end" metadata must be user-friendly, containing data fields and categorizations that CAP students and faculty would recognize.
- The "back end" metadata must conform to internationally recognized metadata standards and be suitable for Open Archives Initiative (OAI) harvesters.
- The online collection must be made available as widely as copyright will allow — so that outside educators and the general public may also utilize the collection.

With the above survey of existing materials, and list of collection goals we began our planning the collection and drafting workflows.

Content Management

The first task of any digital collection is to determine if one should develop or purchase a content management system (CMS) to house it. Fortunately, this decision had already been made: prior to the beginning of this project, **Ball State University Libraries** had purchased **CONTENTdm** to form the base of all collections in our **Digital Media Repository**.

As with any turnkey system, **CONTENTdm** has the disadvantage of already being a fully formed container. Homegrown systems are far more advantageous in this regard, and can be developed with a specific collection in mind for a tailor-made fit. This being said, however, **CONTENTdm** is a surprisingly flexible container, and has the added advantage of being ready to go practically out of the box.

Metadata

With our CMS in hand, we set about determining how to utilize the existing metadata.

As previously stated, the 35mm slides were already cataloged into group-level **MARC** records, with the title and call number of each individual image stored in the 505 field [See sample — Appendix A]. So, some programming was developed to extract the data from

continued on page 34

Appendix A — Sample group-level MARC record.
Art Institute of Chicago (Chicago, Ill.). Grant Park Garden [slide]

000:		: gm5 n FBU
007:		: gs uj jk
008:		: 910522s1987 waunnn 0sneng d
040:		: |cArch
049:		: IBSO
110:	2	: Skidmore, Owings & Merrill.
245:	10	: Art Institute of Chicago (Chicago, Ill.).|pGrant Park Garden|h[slide].
260:		: Seattle, Wash. :|bArt on File,|cc1987.
300:		: slides|bcol.
440:	0	: Place as art : pocket parks and gardens in the city (Series)
500:		: 1977
505:	0	: Art Inst.of Chicago. Grant Park Gdn. Overview: 2097-006 -- Plant bed: 2097-007 -- Arch: 2097-008, 2097-009.
596:		: 2
650:	0	: Gardens.
650:	0	: Urban parks.
650:	0	: Landscape architecture.
856:	40	: |uhttp://libx.bsu.edu/cdmlink.php?ckey=700018&coll=BSU_Arch SlidesCpght|yClick to view available images of this site or work.

Discrete Criteria ...
from page 31

Endotes

1. **Brian Kelly**, "Choosing a Metadata Standard for Resource Discovery." *http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-63/html/*

2. **Yannis Ioannidis**, "Digital Libraries at a Crossroads," *International Journal on Digital Libraries 5, no. 4* (2005): 255-265.

3. **The Getty Museum**. "Metadata Standards Crosswalks" In Introduction to Metadata: Pathways to Digital Information. Online edition, version 2.1, *http://www.getty.edu/research/conducting_research/standards/intrometadata/metadata_element_sets.html*

4. **Dublin Core Metadata Initiative**. Glossary, s.v. "Application profile." *http://dublincore.org/documents/2001/04/12/usageguide/glossary.shtml*

5. **Erik Duval**, **Wayne Hodgins**, **Stuart Sutton**, and **Stuart L. Weibel**, "Metadata Principles and Practicalities," *D-Lib Magazine 8, no. 4* (2002). *http://www.dlib.org/dlib/april02/weibel/04weibel.html*

6. **Lynne C. Howarth**, "Metadata Schemas for Subject Gateways," *International Cataloguing and Bibliographic Control 33, no. 1*, (January/March 2004):8-12.

Zen and the Digital Collection ...
from page 32

the group-level **MARC** record and use it to create the basis of individual records in a **Microsoft Access** database (repeating group level information as necessary for each individual item). Additional programming was developed to draw inferences and to embellish upon the existing data. For instance, once the site "Chicago, Ill." is known, other location thresholds can be created as well: Illinois, Midwest States, United States, North America, etc.

Naturally, even the best programming is not perfect. Portions of this conversion must be assisted by cataloging librarians, and quality assurance testing is an absolute necessity; however, quite a bit of the work can be automated. Thus, through automated scripting and human-intervention, the **MARC** data can be crosswalked and expanded into data fields that are user friendly, and labeled specifically with architecture researchers in mind [See sample — Appendix B].

The resulting **Access** database with "front end" user-friendly data fields will be used to bulk-upload into our CMS once the scanned images are prepared. Once the data is in the CMS, the "front end" metadata set will be tied to a "back end" metadata set of qualified **Dublin Core** elements. In this manner, our "front end" users may utilize the collection with the highly granular and user-friendly data fields, while simultaneously, our collection will offer OAI compatible **Dublin Core** records to "back end" users through a wide variety of metadata harvesters, and search engines.

Digital Images

To prepare our images, the architecture librarians first pull all of the individual slides associated with a given **MARC** group record. The slides are bundled and sent to the library's Digitization Center, where they are scanned into high-resolution uncompressed TIFF images and saved with the individual call number as the filename so they can be easily matched to the metadata record. Automatic scripting creates a derivative JPEG from the archived TIFF, suitable for Internet delivery and classroom projection.

Maintaining the copyright protections of the images was perhaps the most difficult obstacle to overcome in building this collection. Largely, the digitized images fell into one of three copyright "levels":

1. Public: These images are public domain images, orphaned works, and/or Ball State University owns the copyright and wishes to share the image with the public at large. Anyone, anywhere, may access these images.
2. Student: Ball State University has paid a licensing fee for the use of these images within the **Ball State** community, or **Ball State University** owns the copyright but wishes to restrict access. These images may only be viewed by students, faculty, and staff.
3. Faculty: **Ball State University** claims no rights to these images, and they may only be utilized by faculty during face-to-face instruction in accordance with the Teach Act. To insure that fair use provisions are not violated, use of these images is monitored, and excessive use spurs either the securing of licensing for the image,

continued on page 36

against the grain **people profile**

Head of Metadata and Digital Initiatives
Ball State University, University Libraries
BL-025, Ball State University, Muncie, IN 47306
Phone: (765) 285-5718 <jabradley2@bsu.edu>

James A. Bradley

BORN & LIVED: Born: Flatwoods, KY. Lived: Bowling Green, KY (udergraduate); Bloomington, IN (graduate); Chicago, IL; Muncie, IN (present).

EARLY LIFE: Raised in very small town in Appalachian area of Eastern Kentucky.

FAMILY: Wife, **Mary**, and one child, **Sylvia**.

EDUCATION: BFA (Painting and Photography) — **Western Kentucky University**; MIS — **Indiana University**.

FIRST JOB: Mowing lawns (12 years old).

FIRST JOB OUT OF GRAD SCHOOL: Structured Text Specialist (XML) for the **American Theological Library Association** in Chicago, IL.

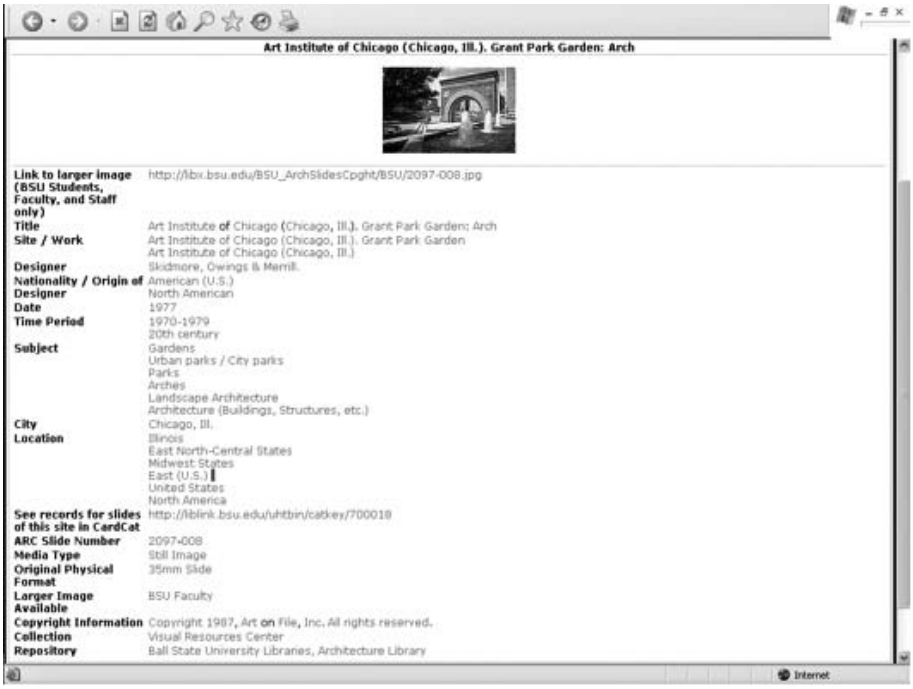
IN MY SPARE TIME I LIKE TO: Spend time with my family, play with my daughter (2 years old this February), visit extended family in Kentucky, visit my former neighborhood in Chicago (Wrigleyville/North Center — GO CUBS!).

FAVORITE BOOKS: **Roald Dahl**, **Patrick Suskind**, **Stephen Dobyns**, **Kurt Vonnegut**...

PHILOSOPHY: "Here we are, trapped in the amber of this moment. There is no why." (The quote is from **Kurt Vonnegut**'s novel, *Slaughterhouse Five*. Jumped into mind when thinking about my favorite books.)

HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS: An increasing role of digital assets and increasing interconnectivity. Interconnectivity being the key, we still have far too many untapped resources and "data silos" which must be unlocked and shared.

Appendix B — Sample CONTENTdm record



Zen and the Digital Collection ...
from page 34

or (as a final alternative) the removal of the image from the collection.

Our CMS, **CONTENTdm**, does possess some internal methods to restrict access on a collection-level or item-level basis; however, it does not allow for multiple levels of access to a single collection. Complying with these copyright provisions was not optional, it simply must occur. If we could not find a way to stretch the container, the collection would have to be divided into multiple smaller collections.

Fortunately, after much head scratching, we developed an architecture that would allow a unified **CONTENTdm** collection to offer materials to a variety of user types while still protecting image copyright. The solution presented itself when our research determined that the sharing of thumbnail images within an Internet search engine is perfectly legal (see ***Kelly v. Arriba Soft***, a.k.a. **Ditto.com**).

Thus, we could load the metadata into **CONTENTdm** alongside a thumbnail-sized image. The **CONTENTdm** record would then link out to the larger JPEG images which would reside on a separate file server, and access would be approved or refused based on login using the server controls. So, regardless of copyright, all users may access the metadata record and the thumbnail image; however, only certain types of users can follow the link to the full size image.

Workflow And Personnel

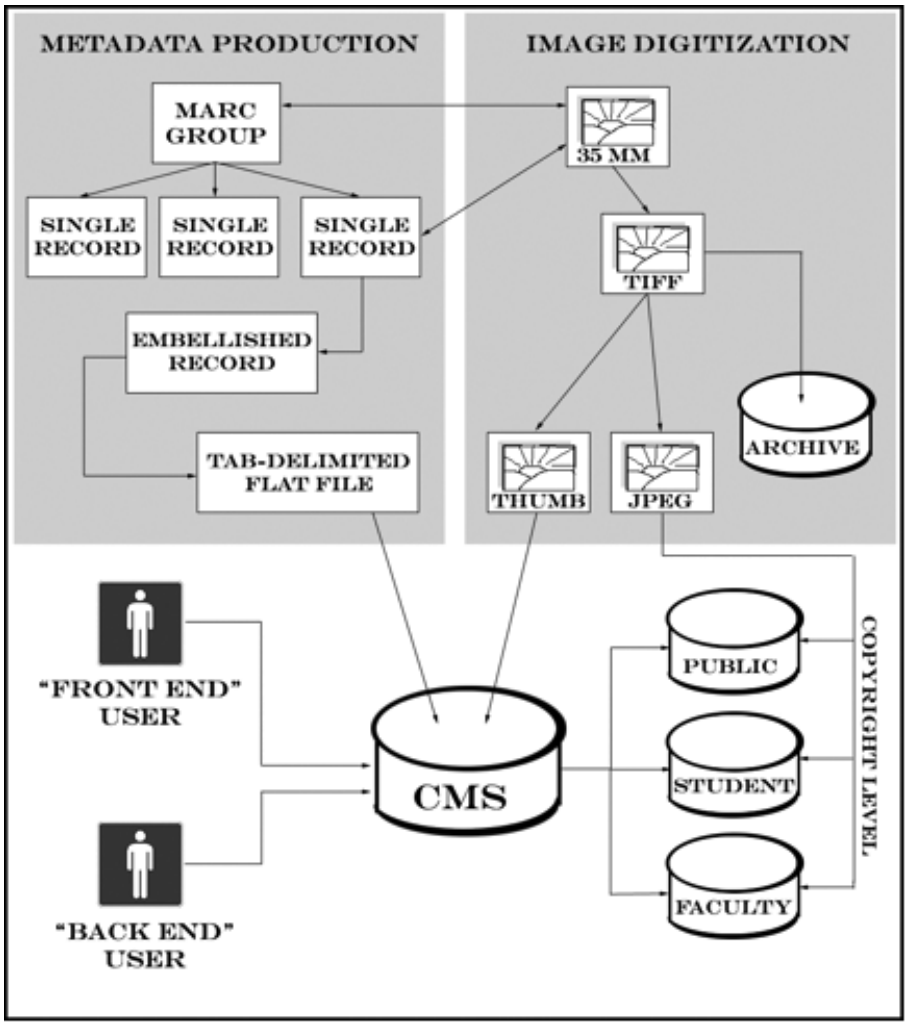
The workflow of the collection (See Diagram — Appendix C) is fairly straightforward once you understand the reasoning behind the parts.

The Metadata Team works to extract data from the group-level **MARC** files, embellish it, and then export the final result to a tab-delimited flat file for bulk loading into the CMS. Likewise, the Digitization Team scans the original slide and archives the TIFF version. Automatic scripting takes over from there to produce thumbnail and medium resolution JPG derivatives. Finally, the local call number acts as a data point that allows for the reunion of metadata record to the correct scanned image during the upload process.

The ability to have metadata production proceeding simultaneously, and independently of image digitization has been crucial to the success of this project. Compartmentalizing the work has allowed us to create our digital collections with very few new personnel. Student employees handle most of the scanning, and existing personnel from various departments (Cataloging and Metadata, Physical Processing, Archives & Special Collections, and Library Information Technology Services) have been drawn in according to need, specialty, and ability to complete portions of the workflow.

Additionally, the ability to automate many tasks within the process has allowed us to move forward with surprising speed. We are eighteen months into the project and our current collection is nearing 40,000 records — an average of five hundred records per week.

Appendix C — Workflow Diagram



Conclusion

We at **Ball State University Libraries** invite you to view our **Architecture Image Collection**, and any of the other collections residing in our **Digital Media Repository** (*http://libx.bsu.edu*). Each collection has brought with it a unique set of challenges and obstacles to be overcome.

When creating a digital collection, sometimes, you can bend the object. On other occasions, you can stretch the container. Most often, however, you will find you need to do a little bit of both.

Rumors
from page 26

chapters choc-a-block full of important legal information. It's on my beside table to read. Check it out!

Speaking of which, Audrey Fenner has edited a book, ***Integrating Print and Digital Resources in Library Collections*** (The Acquisitions Librarian, no. 35/.36, 2006). See our review, this issue p.54.

Speaking of reading, have been fascinated by the recent book, ***The Man Time Forgot, A Tale of Genius, Betrayal, and the Creation of Time magazine*** by **Isaiah Wilner** (HarperCollins, 2006). It's all about **Briton Hadden**, the man who with **Henry R. Luce**, began *Time* magazine when they were students at **Yale**. **Hadden** died at the age of 31.

And this news from the alert **Chuck Hamaker** <cahamake@email.uncc.edu> — From **Times online** (15 February 2007) and an

continued on page 40